Home : Tutorials : XML Tutorial

by **Shermin Voshmgir** © 1999

Introduction

1986:	SGML ISO 8879-1986
Nov 1995:	HTML 2.0
Nov 1996:	Simplified and stripped down SGML draft (dubbed XML)
Jan 1997:	HTML 3.2
Aug 1997:	XML working draft
Dec 1997:	XML 1.0 proposed recommendation HTML 4.0 Recommendation
Jan 1998:	XML
Feb 1999:	XHTML

SGML

In 1986 the Standard Generalized Markup Language (**SGML**) became an international standard for defining descriptions of the structure and content of different types of electronic documents. SGML, the "mother tongue" of HTML and XML, is used for describing thousands of different document types in many fields of human activity, from transcription of ancient Sumerian tablets to the technical documentation for stealbombers, and from patient's clinical records to musical notations.

SGML has withstood the test of time. Its popularity is rapidly increasing among organizations with large amounts of document data to create, manage, and distribute as in the Defense, Aerospace, Semiconductor and Publishing industries. However, various barriers exist to delivering SGML over the Web. These barriers include the lack of widely supported stylesheets, complex and unstable software because of SGML's broad and powerful options, and obstacles to interchange of SGML data because of varying levels of SGML compliance among SGML software packages.

These difficulties have condemned SGML to being a successful niche technique rather than a mainstream tool. Indeed some cynics have renamed SGML in 'Sounds Good Maybe Later'.

HTML

HTML is a subset of SGML, the most frequently used document type in the Web. It defines a single, fixed type of document with markup that lets you describe a common class of simple office style report, with headings, paragraphs, lists, illustrations, etc., and some provision for hypertext and multimedia.

HTML was defined to allow the transfer, display and linking of documents over the internet and is the key enabling technology for the WWW. Prior to the emerging of the

internet, it was unusual in the word of computing to hear the word "page" used to describe elements of data. But HTML web pages have amazing similarities with paper in their role of information publishing. Both HTML and paper pages

- are optimized for visual clarity,
- focus on ultimate usability (but not on reusability),
- contain no contextual information, and
- have no document structure to enable automation.

Today's web is created by Hand for Eyes only. HTML has too low an "Information IQ" to enable many desirable applications. HTML was designed as a markup language an with simple structures, strong emphasis on formatting and was weak for encoding content. It was not designed to encode structure and semantics needed for complex applications.

XML

Because of the lack of SGML support in mainstream Web browsers, most applications that deliver SGML information over the Web convert the SGML to HTML. This down-translation removes much of the intelligence of the original SGML information. That lost intelligence virtually eliminates information flexibility and poses a significant barrier to reuse, interchange, and automation.

For this reason, XML (Extensible Markup Language) was developed by the XML working group (known as the SGML Editorial Review Board) formed under the auspices of the **World Wide Web Consortium (W3C)** in 1996. XML is a highly functional subset of SGML. The purpose of XML is to specify an SGML subset that works very well for delivering SGML information over the Web. When the mainstream Web browsers support XML, it is believed that it's going to be very easy to publish SGML information on the Web. It's actually misnamed because XML is not a single Markup Language. It is a metalanguage to let users design their own markup language.

XML is a public format and not a proprietary format of any company. The v 1.0 specifications was accepted by the W3C as Recommendation on February 10, 1998.

XML was conceived as a means of regaining the power and flexibility of SGML without most of its complexity. While retaining the beneficial features of SGML, XML removes many of the more complex features of SGML that make the authoring and design of suitable software both difficult and costly. But XML also lacks some important capabilities of SGML that primarily affect document creation, not document delivery. That's because XML was not designed to replace SGML in every respect.

The question that is open is not whether XML will succeed as a widespread data format, but rather how fast, to what level of success and with what products. The question of whether XML would enter the market was answered when Microsoft, Adobe, Netscape and other big market players not only supported the development of the new standard but began making sizable product investments to this new format. The leading Web browser Products already support XML in their latest releases. The momentum building behind the XML effort means that XML is inevitably destined to become the mainstream technology for powering broadly functional and highly valuable business applications on the Internet, intranets, and extranets.

XHTML

XHTML is a working draft for there formulation of HTML 4.0 [HTML] as an application of XML 1.0 [XML]. It is the basis for a family of future document types that extend and subset HTML.

There are two major reasons for content developers to adopt XHTML. First, XHTML is designed to be extensible (Design you own tags). Second, XHTML is designed for portability. There will be increasing use of non-desktop user agents to access Internet documents. Some estimates indicate that by the year 2002, 75% of Internet document viewing will be carried out on these alternate platforms. In most cases these platforms will not have the computing power of a desktop platform, and will not be designed to accommodate ill-formed HTML as current user agents tend to do. Indeed if these user agents do not receive well-formed XHTML, they may simply not display the document.

For further information on XHTML read the **XHTML working draft**.

Home : Tutorials : XML Tutorial